EECS 496: Data Science Seminar
Casey Grage & Renee Zha
Group 10: Final Report

Demographics in the Chicago Police Force

Theme: It is often explored whether certain identity groups of civilians face a greater percentage of police violence, or whether certain identity groups of officers face a greater percentage of allegations. But what are the most common pairings? Who is a black male officer most likely to receive allegations from? Who is mostly likely to assert police harassment against a Latina woman? We want to explore a number of these relationships. In other words, we aim to examine if there are trends in allegations/complaints between identity groups of civilians, identity groups of officers, and relationships between identity groups of officers and civilians. We are using “identity group” to indicate a group of people that have the same race and gender. We will also look at location of the allegation and age for certain inquiries.

Checkpoint 1:

For checkpoint 1 we wanted to answer:
1. Which identity groups filed the most complaints (for officer & civilian complainants)?
2. What percentage of unique officers have multiple allegations of any kind against them?
3. Which identity groups are most often victims (for officer & civilian complainants)?
4. What percentage of unique officers have multiple victim allegations against them?

We used simple SQL queries to look at the percentage of both allegations by race and victims by race. We used the data_allegations table to answer #2 and #4.

We found that black individuals filed the most overall complaints against CPDB officers. They were also the most common victims in complaint reports. Further, 7.6% of officers with allegations against them had multiple, making up 59.4% of total force. 89.9% of officers with allegations that included a victim had multiple, which makes up 31.6% of total force.

In future analysis, we would want to answer: how does the racial breakdown and/or likelihood of officers to have multiple allegations vary with officer age?

Checkpoint 2:

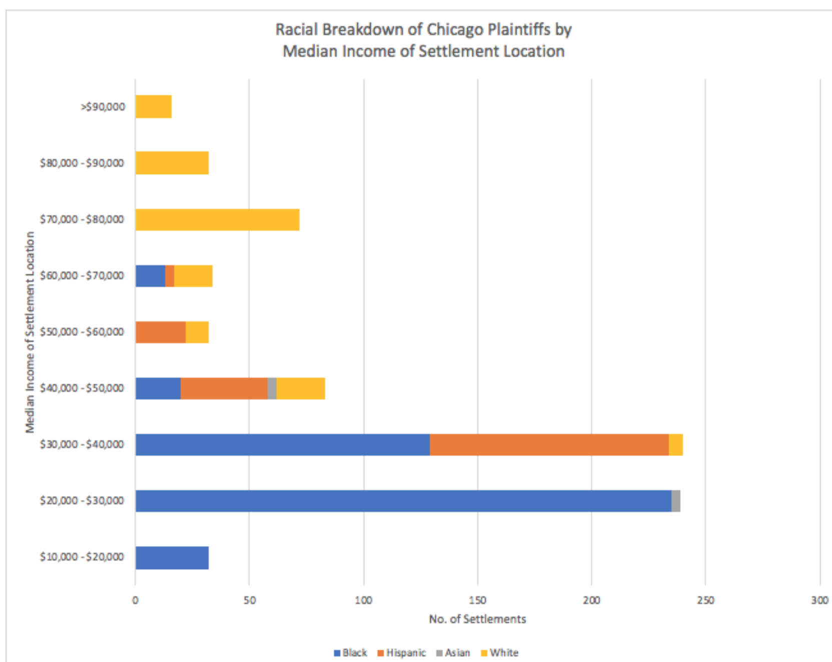
For checkpoint 2, we wanted to answer:
1. Are the identity groups most represented in plaintiffs from the settlement data also most likely to be victims in the complaint reports?
2. Are certain identity groups overrepresented in the complaints and settlement database compared to the general population?
3. Do police officers of certain identity groups have more allegations against them in low income areas than in higher income areas?
4. Are certain identity groups overrepresented as complainants in low income areas verses higher income areas compared to the general population?

We used Trifacta to integrate data from the settlement database with data from the chicago police database. We intended to combine demographic data about civilians from the settlement and cpdb datasets in a pipeline similar to the following:

- Table of total no. of plaintiffs in each identity group: Identify identity groups of plaintiffs in the settlement data using demographic information from cpdb data
- Table of no. of victims in each identity group: Identify most common identity groups of CPDB victims

We were unable to find sufficient identifying or demographic information about the plaintiffs in the provided core settlement database that was published on Canvas, so we had to improvise methods of collecting this data. We mapped settlement location (lat and lon points) to the corresponding CPDB beat_id using the given polygons. We then found the majority race of each beat_id, and assigned that race to the race of the plaintiff. This allowed us to find the percentage of each race in total settlements. We also looked up the median income of each beat_id and used that data to look at trends between race, income_area, and number of settlements. This method also normalizes the number of allegations from people of a given identity group in a neighborhood by their representation in that area's population. Otherwise it is difficult to distinguish whether one group is represented among the plaintiffs.

There doesn't seem to be a huge difference between complainant racial breakdown and plaintiff racial breakdown. The main notable change was that the percentage of black plaintiffs is higher than the percentage of black complainants, while the opposite is true of hispanics. It is possible that black civilians are more likely to file more serious allegations than other races. It is also possible that hispanics are less likely to follow up complaints with legal action. This could be due to obstacles against them. For example, lawyers may be less likely to take their cases.



The majority of settlements happen in lower income neighborhoods. The racial breakdown of these neighborhoods' populations is such that black plaintiffs are by far the most common in low income neighborhoods, while white plaintiffs are the most common in wealthier neighborhoods. The majority of hispanic plaintiffs have settlements that occur in middle to low income areas. It is also interesting to see that the only income area in which all four major races are represented is the lower middle class $40,000 - $50,000. This graph shows that plaintiffs are largely lower income, and largely black and hispanic. However, the number of plaintiffs steadies for the middle to upper classes.

Because of the limited plaintiff demographic information, we were unable to analyze the overall demographics of plaintiffs in both the settlement and cpdb datasets. However, to gain some initial insight into this problem, we found the percent chance that any given race will be a complainant (by comparing number of complainants by race compared to their total Chicago population) using the data_racepopulation table from cpdb.

From this initial exploration, we see that there is a wide range in the percentage of people in a certain identity group who file complaints. In particular, the rate of complaints from people who are black is almost triple that of those who are white, and 12x the rate of those who are Asian. This suggests that there may be underlying factors (cultural, societal, biases, etc.) that affect these rates.

Then, we used Trifacta Wrangler to join the data_officerallegations table with information from data_officer, data_allegation, and data_area, in order to combine demographic information about the listed officers with the areas (beats) that the allegations took place.
We used this site to find the neighborhoods with the highest and lowest average incomes.
We used this site to find the beats within neighborhoods of interest.
We used this site to find racial demographics of neighborhoods of interest.

In higher income neighborhoods, black officers were *more* likely to have allegations against them compared to the overall population of black people while white officers were *less* likely to have allegations against them compared to the general population in that neighborhood. Interestingly, Asian officers were *much* less likely to have allegations against them compared to the overall population, but this could be due simply to the small sample size. In lower income neighborhoods, white officers were much more likely to have allegations against them compared to the general population while black officers were much less likely to have allegations against them compared to the general population. From this, we learned that income area matters a lot in correlating whether an officer of a certain race is more or less likely to have an allegation against them compared to the general population.

We used Trifacta Wrangler to join the data_complainants table with information from data_allegation and data_area to combine demographic information about the people who filed complaints with the areas (beats) that the allegations took place.

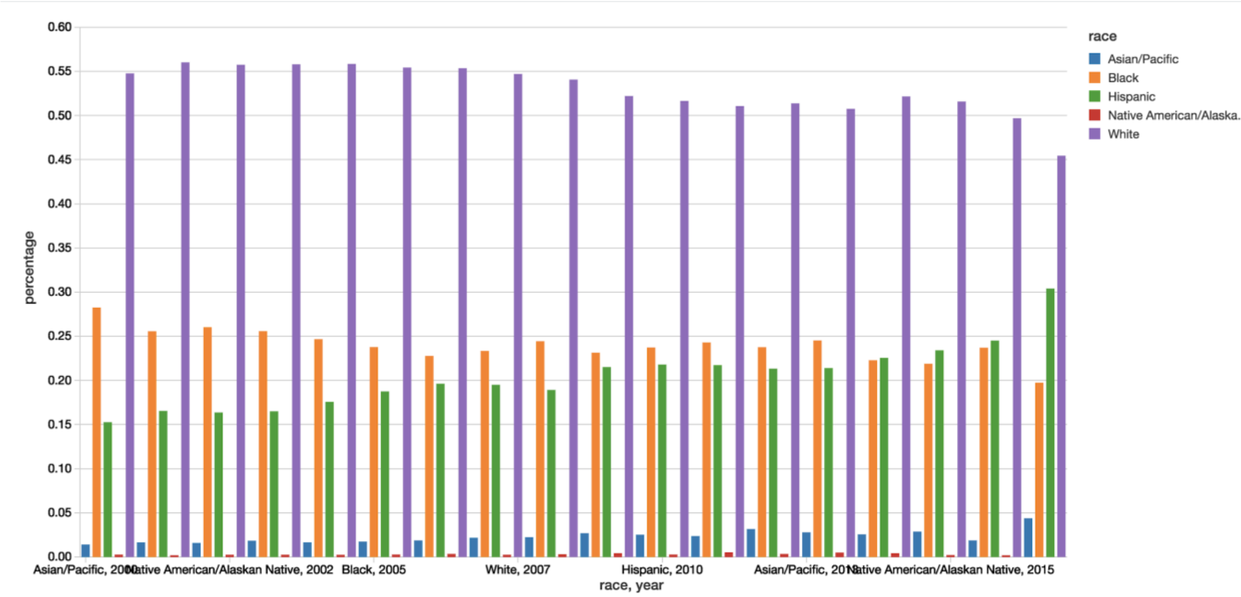
In high income neighborhoods, black civilians were disproportionately *more likely* to file complaints against officers than their reflected populations in the area, in many cases on a scale of 5x to 10x higher than their percentage of the population. On the other hand, White civilians were disproportionately *less likely* to file complaints in proportion to their overall percentage of the population. This may suggest underlying differences in treatment by police officers towards civilians of different races. In lower income areas, this disparity is not as apparent. It is interesting to note that the demographic makeups of these areas are not as “white-dominant,” so it may be difficult to draw wide conclusions among lower-income areas. Despite this, in general, black civilians remain the most likely group to file complaints against officers, suggesting a common pattern.

Checkpoint 3:

For Checkpoint 3, we wanted to answer:

1. Are the percentages of allegations against officers of any identity group rising or falling compared to the rates of total allegations *and* to the rates of population demographics in Chicago?
2. Are investigators who are in the same identity group as the officers named in the complaints less likely to take action against the officer?
3. Are investigators who are in the same identity group as the victims/complainants more likely to take action against the officer?

We used Apache Spark in Databricks to easily visualize the data from SQL queries. *[Here is the databricks link if desired.]*



We found the percentage of allegations against native americans, asian/pacific islanders, and white officer look relatively similar to the actual racial breakdown of the Chicago Police Department. Also, all of these numbers have stayed relatively constant from 2000 to 2017, except that the percentage of both total white officers and allegations against white officers has dropped slightly since 2000. However, from 2000 to 2017, the the percentage of hispanic officers in the CPD has risen greatly while the the percentage of black officers has declined greatly. The the percentage of allegations against black officers has declined slightly and the the percentage of allegations against hispanic officers has risen slightly, however these changes do not exactly imitate the changes in overall CPD population demographic. This means that black officers are more likely to have allegations against them than hispanic officers, and this reality has become increasingly true from 2000 to 2017.

We found black officers were more likely to be disciplined by black investigators than white officers by white investigators and hispanic officers by hispanic investigators. In future analysis, we would want to know: how significant is this difference? How does the percentage of each of the other investigator-officer race pairings compare?

White victims and complainants are far more likely to result in disciplinary action taken against the officer by a white investigator compared to black or hispanic victims and complainants. This finding confirms our suspicion that white civilians may have a louder voice when it comes to police misconduct.

<u>Checkpoint 4:</u>

For Checkpoint 4, we wanted to answer:

1. Given data from years ‘A’ through ‘E,’ can we predict how the rate of complaints that result in disciplinary action against the officer will change over time for years ‘F’ through ‘H’? Does this model get better or worse with added features (in addition to year): officer gender, officer race, complainant race, and complainant gender?

2. Given an officer, and the nature of the complaint (TRR details), can we predict the subject identities? What are the most important features?
3. Given a subject, and the nature of the complaint (TRR details), can we predict the officer identities? What are the most important features?

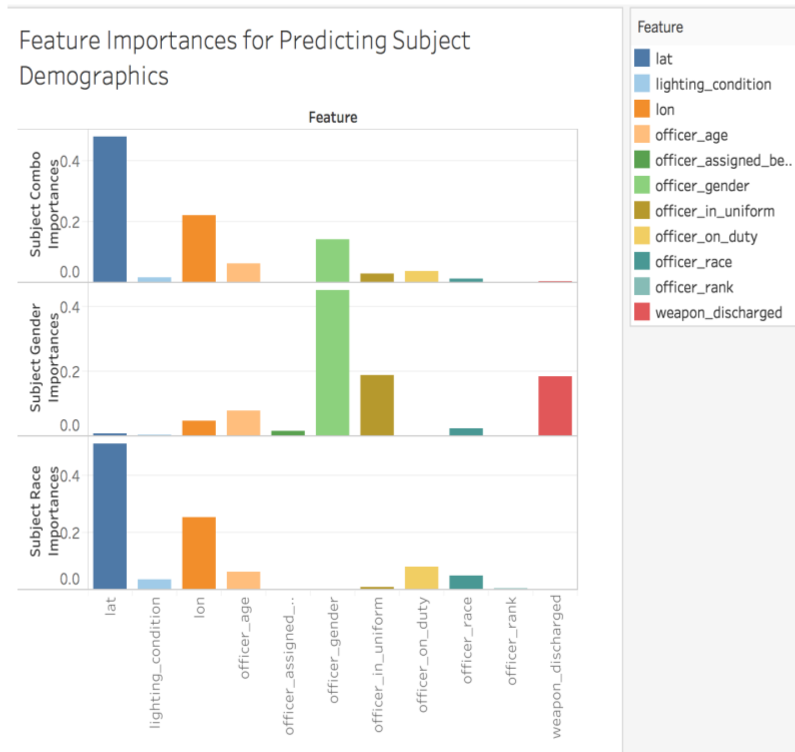
We used Databricks to get spreadsheets of all of the desired CPDB data using SQL queries. *Here is the databricks link if desired.* We then used Excel to make the calculations. We esed years 2006-2011 to train — except for White M officer / White F complainant, we used 2007-2011 due to limited data. We used years 2012-2015 to test. We used *AVERAGE()* on the databricks exports to get training and testing averages, *SLOPE()* to get training slope, *(5*training slope)+training avg.* to get predicted average (5 is the average change in years between training and testing), and *ABS(predict avg. - actual avg.)/actual avg.* to get the mean relative error.

Officer Ra	Office	Complainant	Compl	Training Average	Training Slope	Predicted Avg.	Actual Avg.	Mean Relative Error
Black				8.4912	-0.8417075	4.2826625	10.2594	0.582562089
White				4.4523	-0.0634228	4.135186	4.9368	0.162375223
		Black		2.7197	-0.4288327	0.5755365	3.0391	0.810622717
		White		18.0783	-0.341427	16.371165	18.233	0.102113476
		Black	M	2.8645	-0.4048764	0.840118	3.0675	0.726122901
		White	M	21.0564	-0.6295023	17.9088885	20.5676	0.129266978
		Black	F	2.6124	-0.4605658	0.309571	3.025	0.897662479
		White	F	11.8178	0.73055228	15.4705614	13.8211	0.119343714
White	M	White	M	16.6488	0.09618	17.1297	15.8281	0.082233496
White	M	Black	M	2.1838	-0.18227	1.27245	1.7995	0.292886913
White	M	White	F	10.479	0.50476	13.0028	10.2015	0.274596873
White	M	Black	F	1.0985	0.06927	1.44485	1.9839	0.271712284
Black	M	White	M	31.7758	-2.2252	20.6498	32.3231	0.361144197
Black	M	Black	M	3.7397	-0.8417	-0.4688	5.3899	1.086977495
Black	M	Black	F	4.4522	-0.2488	3.2082	4.82	0.33439834
Black	M	White	F			*not enough data*		
						Avg. Errors Given...		
						Officer Race		0.372468656
						Complainant Race		0.456368096
						Complainant Race & Gender		0.468099018
						All Data		0.386278514

There is no difference in the accuracy of models when training sets vary based on demographic data. In other words, adding more features does not improve the linear regression model. This is because there aren't consistent trends for any of the combos. This means that rates of complaints successfully resulting in disciplinary action, while they do vary with complainant and officer demographics, have not changed with any significant trend in the last 10+ years. We would like to know, would using more years as training/testing data improve the model?

We used Databricks to create random forest classifiers with the desired features from the CPDB using python. *Here is the databricks link if desired.*

First, we chose the features we would like to analyze as predictors of subject/officer identities. We used SQL terminal to get data, then manipulated that data and joined tables in databricks to get the exact dataframe we want to work with. Then we turned some variables into booleans when we only cared about yes/no results. For all other variables, we used one-hot encoding to allow us to work with categorical data. We created six random forest classifiers to predict 1) subject gender, 2) subject race, 3) subject gender/race combination, 4) officer gender, 5) officer race, and 6) officer gender/race combination.



In descending order, the three most important features for predicting subject gender are officer_gender, officer_in_uniform, and weapon_discharged. It seems male subjects are more likely to have TRRs with male officers, and females with females. Also, if a weapon was discharged, the subject is more likely to be a male. My theory with the officer_in_uniform feature importance, is that officers are very often in uniform and the subjects are very often male. So the importance of this feature was overinflated.

In descending order, the four most important features for predicting subject race are location (lat, lon), officer_on_duty, officer_age, and officer_race. Location makes sense because Chicago is so segregated. Whether the officer is on duty or not is also interesting because it shows how many emergency calls there were. Also officer_race is a predictor of subject race. This may be because officers in one area may be more likely to be the same race as the civilians in that area (as per segregation mentioned above). It may also be that white officers are more likely to target black subjects. Officer_age is interesting. Perhaps certain age groups are more likely to have racial biases.

In descending order, the three most important features for predicting subject gender/race combinations are location (lat, lon), officer_gender, and officer_age. For the aggregate subject identity, location and officer demographics are most likely predictors. Location was the number one predictor of subject_race and officer_gender was the number one predictor of subject_gender, so those make sense and have already been discussed. What's interesting is officer race did not make top 3 when predicting subject race and gender combined. It's possible certain age groups are more likely to have racial and gender biases than are race groups to have combined racial and gender biases.

In descending order, the three most important features for predicting officer gender are subject_gender, officer_in_uniform, and officer_assigned_beat. It's really interesting that officer_rank isn't very important at all in determining gender! The genders correlate because again TRRs are more likely to involve officers/subjects of the same gender. Also, it seems whether the officer is in uniform or in their assigned beat correlates with gender. I think male officers are more likely to get called for / take emergency calls when they're not on duty.

In descending order, the three most important features for predicting officer race are location (lat, lon), subject_race, and officer_rank. Again, Chicago is highly segregated so it's possible officers are probably more likely to be the race of their assigned beat (or they're own neighborhood if they took an emergency call). It's also possible certain officers have racial biases against subjects. Officer rank is interesting. Perhaps there is racial bias in the Chicago Police force or men of color are less likely to have higher education and therefore less likely to be promoted in CPD.

In descending order, the three most important features for predicting officer gender/race combinations are location (lat, lon), subject_gender, and subject_race. These are the three features I would expect to be most important given we know location is correlated with race, races are correlated, and genders are correlated.

Checkpoint 5:

For Checkpoint 5, we wanted to predict the identity group of an officer given a TRR or complaint report (or the identity group of the victim if applicable) by training the model on the text of complaints and TRRs from different identity groups. We used Tensorflow as a library within Databricks. [*Here is the databricks link if desired*.]()

For the corpus of text, we decided to use the "narratives" column from the cases_case table in the settlements database. These short descriptions provide a story for each case from a seemingly journalistic point of view, making them ideal for performing text analysis on. However, the sample size is not very large (only around ~900 total cases).

We implemented a multi-layer perceptron (MLP) model using unigram and bigram tokenizations to represent each narrative (with heavy guidance from [this resource]() to solve this classification problem. Because of the small sample size, we decided to classify only based on the officer's race, and not by their gender as well. Doing so would have probably resulted in too few representatives in certain categories (American Indian and Asian/Pacific females especially), though the distribution of race itself is already quite unbalanced and likely affected the model's accuracy.

After attempts to tune hyperparameters (4 32-unit intermediate layers):

```
Train on 2345 samples, validate on 261 samples
Epoch 1/1000
 - 5s - loss: 1.5910 - acc: 0.4725 - val_loss: 1.5588 - val_acc: 0.5556
Epoch 2/1000
 - 2s - loss: 1.5066 - acc: 0.5753 - val_loss: 1.4301 - val_acc: 0.5556
Epoch 3/1000
 - 2s - loss: 1.3030 - acc: 0.5795 - val_loss: 1.1679 - val_acc: 0.5556
Epoch 4/1000
 - 2s - loss: 1.0886 - acc: 0.5872 - val_loss: 1.0919 - val_acc: 0.5556
Epoch 5/1000
 - 2s - loss: 1.0060 - acc: 0.6000 - val_loss: 1.0392 - val_acc: 0.5594
Epoch 6/1000
 - 2s - loss: 0.9208 - acc: 0.6307 - val_loss: 1.0047 - val_acc: 0.5900
Epoch 7/1000
 - 2s - loss: 0.8527 - acc: 0.6742 - val_loss: 0.9951 - val_acc: 0.6207
Epoch 8/1000
 - 2s - loss: 0.8071 - acc: 0.6968 - val_loss: 1.0063 - val_acc: 0.6169
Epoch 9/1000
 - 1s - loss: 0.7690 - acc: 0.7126 - val_loss: 1.0146 - val_acc: 0.6245

Model evaluation:

 32/652 [>.............................] - ETA: 0s
192/652 [=======>......................] - ETA: 0s
384/652 [================>.............] - ETA: 0s
640/652 [============================>.] - ETA: 0s
652/652 [==============================] - 0s 265us/step
Accuracy: 0.596625766505
```

Despite changing around parameters including # of layers, # of units, and learning rate of the model, the test accuracy seemed to stay around 59-60%. Considering that the percentage of white officers in this particular dataset is also around 58%, it is tough to say that this model is much better at predicting an officer's race from just guessing based on demographic populations. Again, most likely limitations are due to the fact that the corpus of data was not very large, and more investigation needs to be done with other sources of text data (complaint reports?). In addition, more experimentation could be performed on different types of models (the MLP model was suggested based on the metrics of the data, however models that rely on sequence tokenizations could also be investigated). In particular, it would be interesting to see if the “sentiment” of the narratives change substantially based on the identity groups of the people involved.

Checkpoint 6:

For Checkpoint 6, we wanted to answer:

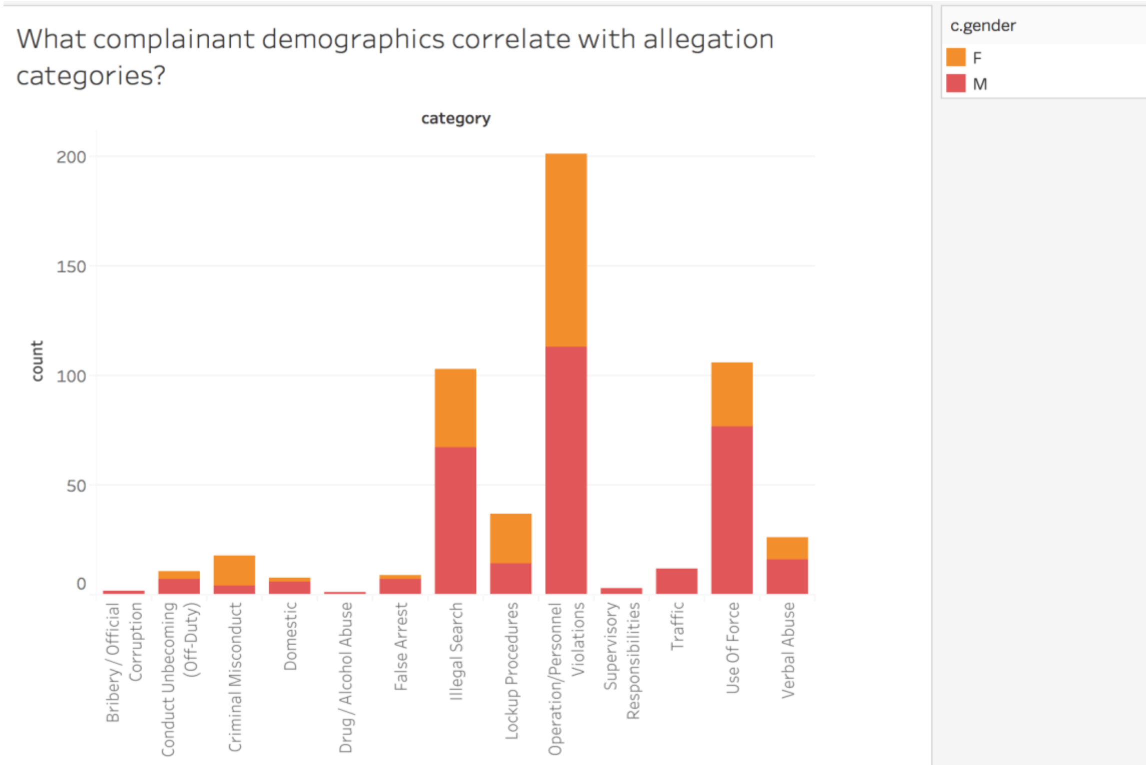
1. Which pairings between identity groups of officers and their respective accusers are most common?
2. Which identity groups make an officer most likely to have an allegation?
3. What identity groups of civilians make them more likely to file complaints/accusations for each type of harassment?
4. What identity groups of officers make them more likely to have complaints/accusations for each type of harassment filed against them?

We used SQL queries in terminal to gather the data, then we imported that data into Tableau to better understand and visualize trends. In general, we found that bar graphs and stacked bar charts were the most useful methods of visualization. It was also helpful to stack bar charts on top of one another (not stacked). For example, we often showed the racial breakdown of CPD overall on one bar graph, and then the racial breakdown of the conditions we were analyzing on a bar graph on top of the demographic one. In another example, one graph displayed allegation likelihood by gender as well as allegation category. We displayed the overall breakdown of gender below this graph to emphasize the gender differences. Sometimes we would also change the color coding of a graph to emphasize different features. In the previous example, we color coded gender to highlight and clarify the likeliness of males/females of different races to have allegations against them.

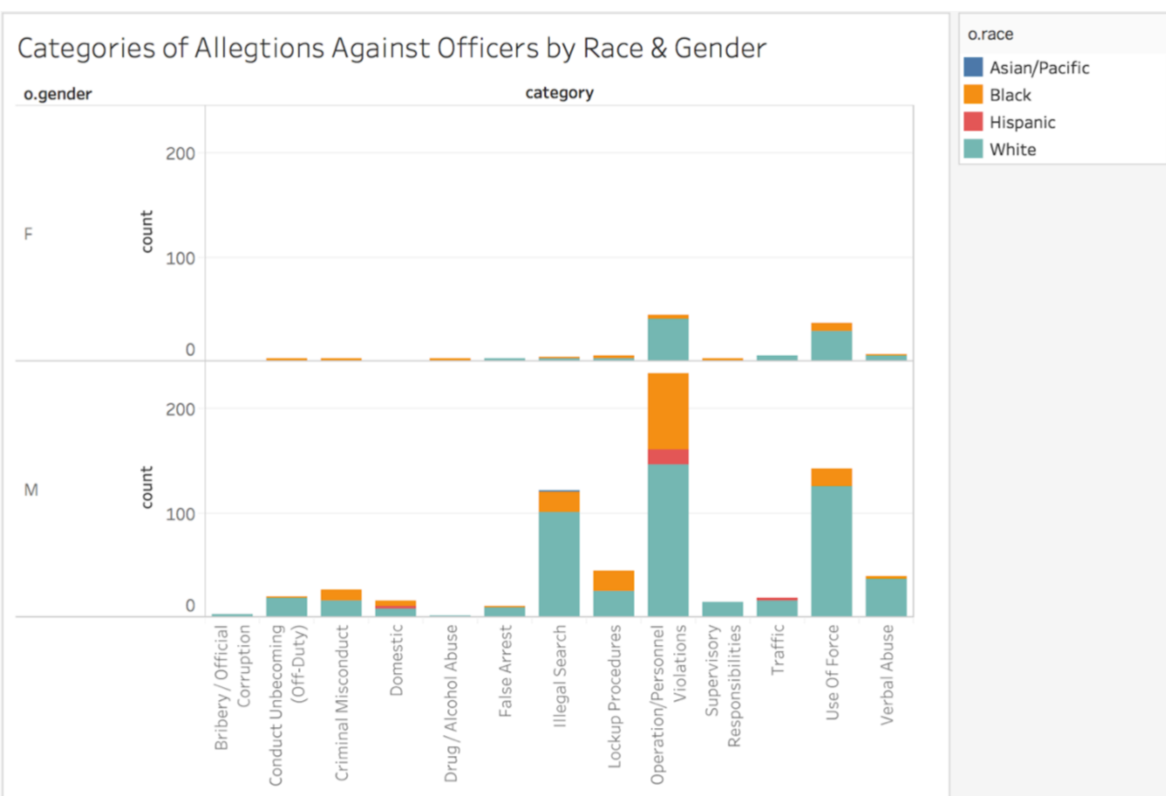
We found that white male officer / black male complainant combinations are by far the most common, with white male officer / black female complainant combo and, interestingly, black male officer / hispanic female complainant combo coming close behind. White male officers are far more likely to have an allegation, then black males, then white females, followed by black females, hispanic males, and asian/pacific islander males.

The vast majority of allegations are for ‘operations/personal violations’ which is nearly 50-50 male/female (which makes sense since Chicago’s population is also 50-50 male/female). Following ‘operations’ in commonness are the categories ‘use of force’ and ‘illegal search,’ for which males file slightly more allegations. Females are more likely than males to file allegations for ‘lockup procedures & criminal misconduct’. White males are much more likely to file allegations than are white females. Hispanic females are much more likely to file allegations than are hispanic males. Black males and hispanic females are the most likely to file allegations. Black males are more likely by far than any other demographic to file an allegation for use of force. ‘Operations/Personal Violation’ is the most common allegation category for every demographic.

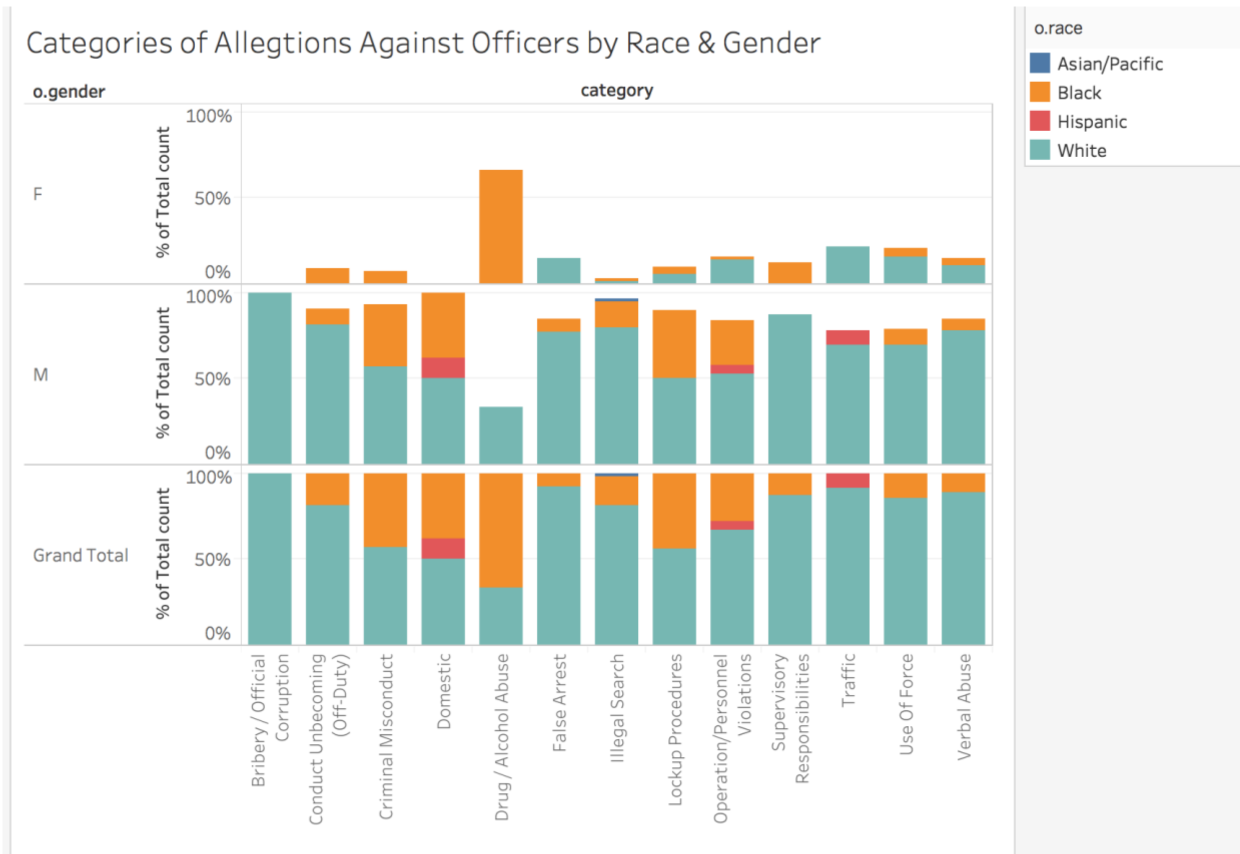
Graph is displayed on the next page.



White officers are the most likely to have allegations against them for everything. Hispanic officers only have a significant number of allegations against them for 'operation/personal violations.' White male officers have significantly more allegations against them for 'use of force,' 'illegal search,' and 'verbal abuse' than is any other demographic. Males overall are far more allegations against them than are females. Of all of the categories, females are most likely to have allegations against them for 'operation/personal violations' or 'use of force'.



One graph we made displaying the percent racial breakdown of each allegation category, shows that all of the 'bribery/official corruption' allegations are against white males. Well over 50% of the 'drug/alcohol abuse' allegations are against black females. White females are actually the most likely to have a 'operation/personal violation' allegation or 'use of force' allegation against them when you compare their counts to the total number of white females in the police force. Black males are second most likely to have an 'operation/personal violation' allegation against them. White males are most likely to have 'illegal search' allegations against them.



Conclusion:

Broadly speaking, findings confirmed suspicions that there were inequalities in how individuals of different identity groups were treated that did not match overall demographics of the populations. More rigorous statistical analysis needs to be conducted to infer stronger correlations.

With respect to machine learning models, location and subject identities are most likely predictors of officer identities and vice versa. However, we had difficulties in building accurate machine learning models to tokenize text. We may need to explore more features in the data.